

LB
1055
B8

UC-NRLF



\$B 71 507

YC 63497

GIF1

LIBRARY OF
UNIVERSITY OF CALIFORNIA

UNIVERSITY OF CALIFORNIA
DEPARTMENT OF EDUCATION

BUREAU OF RESEARCH IN EDUCATION

STUDIES 4, 5, 6, 7.

APPLICATIONS OF PSYCHOLOGY
TO EDUCATION

EDITED BY
J. V. BREITWIESER

4. Modoc County Mental Survey. Frederick J. Adams.
5. False Definition Test in Seventh and Eighth Grades. Adele Bischoff.
6. Training for Rapid Reading. J. V. Breitwieser.
7. A Study of Individual Retests. Elise H. Martens.

MAY, 1922
PRICE 25 CENTS

UNIVERSITY OF CALIFORNIA
DEPARTMENT OF EDUCATION

BUREAU OF RESEARCH STUDIES

The Bureau of Research of the Department of Education of the University of California publishes from time to time the results of investigations, discussions of educational issues, and similar matters. These Studies are for sale at indicated prices by the University of California Press, Berkeley, California. Payment must accompany the order if price is One Dollar or less.

1. Mead, Cyrus D. Measuring Class Room Products in Berkeley..... \$.50
2. Hart, Frank W. A School Building Survey and Schoolhousing Program
for Napa, California..... .50
3. Palmer, Emily G. A Survey of the Garment Trades in San Francisco.... .40
4. Adams, Frederick J. Modoc County Mental Survey.
5. Bischoff, Adele. False Definition Test in Seventh and Eighth Grades.
6. Breitweiser, J. V. Training for Rapid Reading.
7. Martens, Elise H. A Study of Individual Retests.
Nos. 4, 5, 6, and 7 in one cover..... .25
8. Hart, Frank W., and Peterson, L. H. A School Building Survey and
Schoolhouse Program for San Rafael, California..... .50

UNIVERSITY OF CALIFORNIA
DEPARTMENT OF EDUCATION

BUREAU OF RESEARCH IN EDUCATION

STUDIES 4, 5, 6, 7

APPLICATIONS OF PSYCHOLOGY
TO EDUCATION

Edited by

J. V. BREITWIESER

Associate Professor of Education

CONTENTS

4. MODOC COUNTY MENTAL SURVEY. FREDERICK J. ADAMS.....	5
5. FALSE DEFINITION TEST IN THE SEVENTH AND EIGHTH GRADES. ADELE BISCHOFF	11
6. TRAINING FOR RAPID READING. J. V. BREITWIESER.....	16
7. A STUDY OF INDIVIDUAL RETESTS. ELISE H. MARTENS.....	21

TO THE
LIBRARY OF
THE UNIVERSITY OF CALIFORNIA

LB1055
B8

Of the following papers the "Modoc County Mental Survey" and "False Definition Test in the Seventh and Eighth Grades" are abstracts of more extensive theses written in partial fulfillment of the requirements for the degree of Master of Arts. If any one desires to see the data in detail, or to consult the bibliographies, etc., the original theses are accessible in the Library of the University of California.

MODOC COUNTY MENTAL SURVEY

BY

FREDERICK J. ADAMS

In October, 1921, under the direction of Professor J. V. Breitwieser, Department of Education, University of California, and at the request of the county board of education, a survey was made of all the school children of Modoc County in and above the fourth grade of the grammar schools and in the high schools. This survey is unique in that, so far as the writer has been able to discover, no report has been made through the medium of the various periodicals and publications in the field of education or psychology of any other attempt to apply standardized tests to entire counties of the isolated rural type.

Modoc County is the most northeastern county in California, having an area of 3823 square miles with a population according to the census of 1920 of 1.4 persons per square mile, a decrease of 12.4 per cent in population below the census of 1910. The county is divided into three valleys separated by high ranges; the chief means of communication with the outside world is by way of a narrow gauge railroad running into the most eastern of the valleys, which maintains passenger service three times a week. The chief occupations of the inhabitants of this region are sheep raising, the cultivation of a few orchards, and the growing of a few short-season crops.

The survey applies to children chiefly of American stock, as there are very few foreigners in the county, and the Indian children were not tested. The children were distributed in the schools as follows:

	Course	Teachers	Pupils	
3 Union high schools.....	4 years	5	204	
1 Branch high school.....	2 years	2	19	
1 Branch high school.....	2 years	1	6	
Total high school pupils.....				229
1 Graded grammar school.....	8 grades	7	109	
1 Graded grammar school.....	8 grades	5	70	
Total graded grammar school pupils.....				179
3 Ungraded grammar schools	8 grades	2	62	
35 Ungraded grammar schools	8 grades	1	233	
Total ungraded grammar school pupils.....				295
Total number of pupils tested.....				703

In the grammar schools, the National Intelligence Test was used, being applied to all the pupils in and above the fourth grade. The schools had been in session for a full quarter at the time the tests were given. The National Intelligence Test is divided into two parts, Scale A and Scale B. Each part requires about a half-hour; a recess period of fifteen minutes preceded each part. The examination covers the following kinds of tests:

SCALE A

1. Arithmetical reasoning
2. Sentence completion
3. Logical selection
4. Same-opposites
5. Symbol-digit

SCALE B

1. Computation
2. Information
3. Vocabulary
4. Analogies
5. Comparison

Each of these ten parts is preceded by a practice exercise of the same type of material as the test which follows it, so that the pupil may become acquainted with the requirements and situations to be met.

In the high schools the Army Alpha Test, form 7, was given. This test is familiar to most persons interested in mental testing, and therefore needs no explanation. The test was given according to the instructions in the Army Manual.

Upon the completion of the scoring and the computation of the results of this survey, each grammar school teacher in the county was sent an eight-page booklet giving the norms for the ten parts of the National Intelligence Test; by grades, for the graded and ungraded groups, and the norm for the county as a whole; together with norms obtained in other localities; an analysis of the test used; suggestions as to its significance, and as to methods of making use of the data presented. In addition to this bulletin, each teacher received a table showing the record made by each of the pupils in his school, by total score, and in each of the ten divisions of the test.

The high schools received records of the scores of their pupils, tables of the scores in which their pupils were divided according to the grammar schools from which they had entered high school, a statement of the norms for the county, the norms for each high school, and for other localities, accompanied by a discussion of the significance of the data presented.

The following table shows the results of the National Intelligence Test in comparison with three other groups, by grades, in terms of median total scores for the groups represented:

School grade	Ungraded	County	Graded	Vallejo	Washington	Pittsburg
4	90.10	91.1	93.85	100	145.5	167
5	122.80	134.9	156.80	146	184.5	187
6	159.05	167.7	178.60	180	219.5	224
7	191.85	200.5	216.20	237	248.5	251
8	219.90	225.0	242.80	264	275.5	281

The scores made by the ungraded schools, it is readily seen, are consistently lower than those made by the graded schools, and these in turn are consistently lower than the scores of the other localities considered above.

If we compute the per cent of retardation of the ungraded groups in terms of the average grade increment in test totals for each of the groups we find them to be

53.51% of a grade below the graded schools
175.11% of a grade below Vallejo
199.93% of a grade below Washington
228.89% of a grade below Pittsburg

This table shows clearly the degree of the retardation of the ungraded group in terms of the progress of the other groups: ranging from over half a year behind the graded, to over two and a quarter years behind the Pittsburg group. If the average grade increment in test totals for the ungraded group is taken as a standard, the graded group is 63.43 per cent of a year more advanced than the ungraded group.

A comparison of the two graded grammar schools shows the following differences in median scores:

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Seven teachers.....	93.85	166.05	190.10	222.10	241.20
Five teachers.....	91.70	128.90	149.25	212.90	250.45

Although the number of cases is small, especially in the eighth grade, yet the results seem to show a better classification of pupils according to mental ability in the larger school.

Because of the small number of high school pupils in the county, I shall compare the group as a whole with groups in other localities without attempting to show the variation among the individual

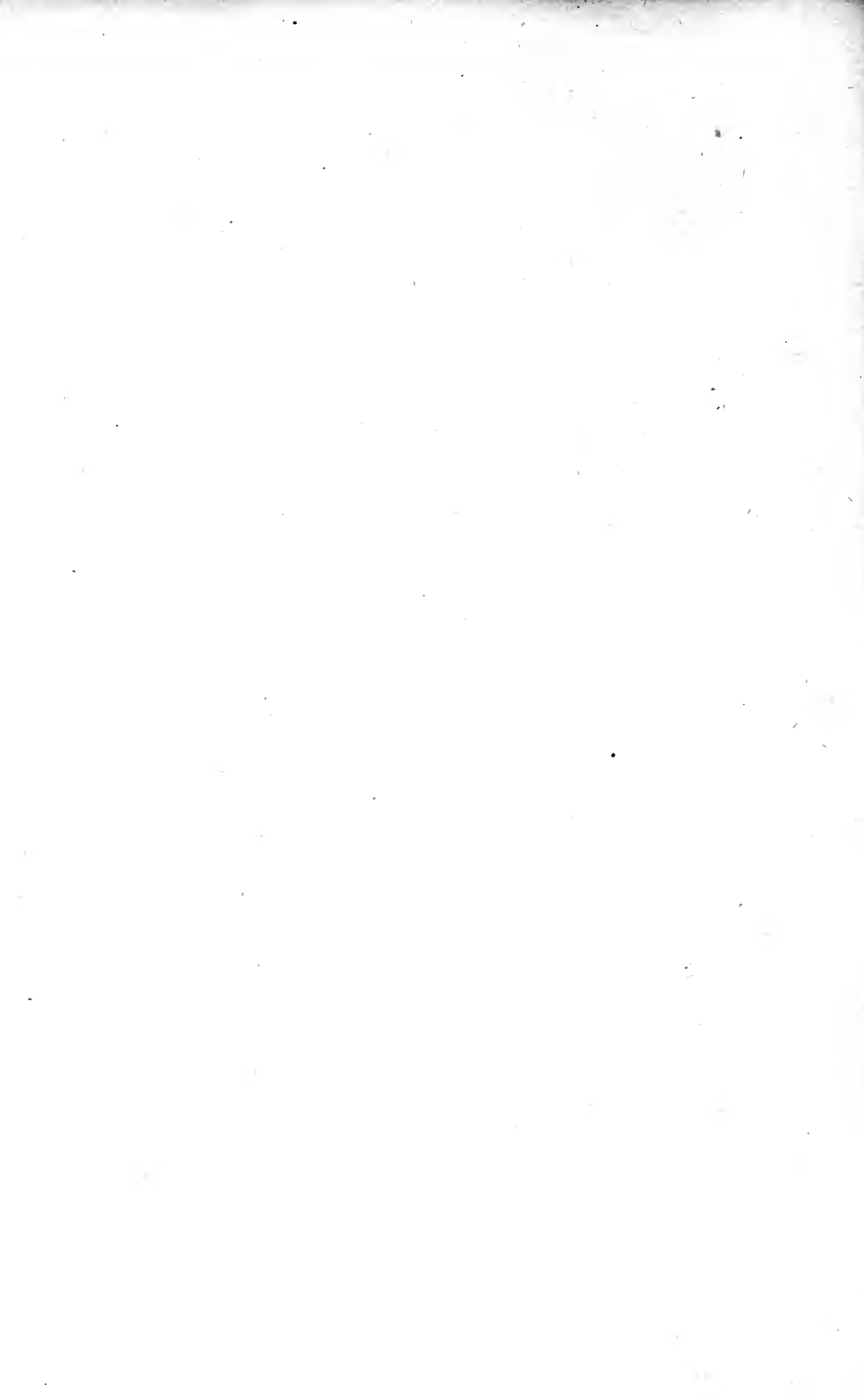
institutions of the county. Measured in terms of medians, the Army Alpha Test scores are as follows, the number of cases being given in parenthesis to the lower right of the score:

	Freshman	Sophomore	Junior	Senior	Graduate
Modoc County.....	77.0 (85)	93.0 (66)	87.0 (45)	106.0 (27)	109.5 (6)
Madison	96.0 (314)	109.7 (296)	122.4 (221)	121.2 (165)	
Rockford	94.0 (500)	107.0 (368)	119.7 (349)	123.8 (260)	
Sioux City.....	107.9 (443)	122.1 (327)	130.2 (250)	137.6 (210)	

The comparison with other localities brings out very clearly the fact that, in terms of the tests applied, students of Modoc County high schools do not have the ability of students in the high schools of the cities listed, the median score for the Modoc seniors being lower than the median score for the Madison and Rockford sophomores, and just a little lower than that of the Sioux City freshmen.

On the basis of the comparison of the results found in the grammar schools of Modoc County with the results found elsewhere by the use of the National Intelligence Test, one might draw one of two conclusions: (a) that the National Intelligence Tests, and perhaps all mental tests, have been standardized and have been formulated *for the use of city children*; or (b) that, in mental ability for their school grades, the children of the grammar schools of Modoc County are below those of other localities. (I might add here that the same and even a greater discrepancy occurs when the age norms are compared with age norms of other localities.) However, when one compares the data on the Army Alpha Test, which was standardized in terms of unselected groups, and upon which norms are obtainable for over two million drafted soldiers, a thoroughly unselected group, we find that the pupils of the Modoc high schools are between two and three years below the high school students in other localities. I believe we must accept to a large degree the second of the above conclusions, i.e., that the children of the schools of Modoc County have less ability, as measured by these tests, than those of the other groups with which they are compared. As we find that the difference is greater for the high school than for the grammar school, so we find that in both the grammar and the high school the difference in ability is increasingly greater. The children of the graded schools show this difference to a

lesser degree than those of the ungraded, and the more nearly standard-graded school shows it to a less degree than the more nearly ungraded school of the two partially graded schools; proving that, although a portion of the difference in norms is due to better school administration, including the selection of pupils, yet the retardation in all the groups of the county shows a consistently lesser ability on the part of the students. Does this not prove, then, the necessity for the reception upon the part of the child of an abundance of stimuli in order to make possible the greatest mental development? And does this not further prove that the poverty of stimuli will lessen the advancement and development of the mental abilities of children, particularly when these are slightly lower than normal at the outset? Here we have fine American boys and girls, not foreigners, suffering mental retardation. In the isolation of rural Modoc County we find that the children, "educated" in the poverty of stimuli of the ungraded school, fail to attain the standards of their more fortunate brothers of the graded school where more time can be spent in enriching their experience and giving to them more stimuli, if only from the words of the teacher and the printed page. How much richer is the experience of the city child, reacting daily to a larger number and a greater complexity of stimuli, thus being enabled to develop his inherent mental ability to a very much greater degree than the child dwelling in the mental poverty of the ungraded school of the isolated community! Is there a greater argument for an enriched curriculum, or for the consolidation of schools than the story told by this survey?



FALSE DEFINITION TEST IN THE SEVENTH AND EIGHTH GRADES

BY

ADELE BISCHOFF

Only within recent years has size of an individual's vocabulary ceased to be merely a matter of speculation and become the subject of scientific investigation. No doubt the growing belief among psychologists that size of vocabulary and general intelligence were to some extent related led to an interest in the subject.

The general and most satisfactory method to determine the vocabularies of children of five years has been to record all the words used by the child during a period of, say, five to eight weeks. Obviously this plan is less feasible for older children since they know a great number of words which they may not use in ordinary conversation. Generally, therefore, the older children have been tested by requiring them to give oral or written definitions of a list of words selected at random from the dictionary.

A decided departure from the foregoing type of vocabulary test is the False Definition Test, formulated in 1915 at Colorado College by Fred M. Gerlach under the direction of Dr. J. V. Breitwieser. They chose a list of one thousand words at regular intervals from the Funk and Wagnall's *New Standard Dictionary*, which contained approximately 375,000 words.

Four hundred of the thousand words which were considered unfamiliar to the average person were placed in a second list. For the remaining six hundred they devised four definitions, of which only one was right and three wrong. The subject was to check the right definition of known words. In the list of four hundred unfamiliar words, the subject was asked to compose definitions of the words which he knew.

The False Definition Test has three decided advantages over other types of vocabulary tests. First, the personal equation in scoring is entirely eliminated, whereas, in a test which requires the subject to devise definitions, quality as well as quantity has to be considered, for definitions are of different grades of correctness. Second, the fact that

a child's expression lags behind his understanding need not be considered. We are all at a disadvantage in defining words, the child doubly so, even although we may understand the ideas they represent. The third advantage lies in the larger size of the dictionary used as the source of the word list. Bonser's¹ tests at Speyer School show an average vocabulary three times the size of that found by Kirkpatrick² for the same grades. Bonser used a dictionary of about 44,000 words while Kirkpatrick used one of 28,000.

In his investigations with college and high school students, Gerlach² determined their average vocabularies in terms of the False Definition Test, and also analyzed the relation of size of individual vocabulary to scholastic status and sex. The present article presents an investigation with the same test upon 155 subjects selected at random from the seventh and eighth grades of the Burbank, Edison, Garfield, and Willard Junior High Schools of Berkeley, California. The subjects comprised approximately an equal number of boys and girls representing an average scholarship of one, two, and three.

To secure the vocabulary index (V. I.) of each individual subject the total number of words having the correct definition checked was ascertained; from this one-third of the number having a wrong definition checked was deducted in order to allow for chance and guessing. The average grade (Av. G.) indicates the average scholarship or mark in class work for the year.

The first four tables, showing an average V.I. of 211.2 for high eighth, 192.7 for low eighth, 164.8 for high seventh, and 152.4 for low seventh, indicate a definite relation between vocabulary and scholastic status. A similar relationship can be seen from the following medians: 218 for high eighth, 191 for low eighth, 160.5 for high seventh and 160 for low seventh. The small difference in the medians of high seventh and low seventh is accounted for by the mean variations of 39.8 and 41.5 respectively.

Although school grades are not always indicative of general intelligence, there is some correlation between them. In this investigation an attempt to correlate vocabulary index with average school grades was made. In each scholastic group the individual indices were divided into three groups according to size of vocabulary range. These three groups were then correlated with the individual average grades. The resulting coefficients of correlation were .48 for high eighth, .54 for low eighth, .81 for high seventh, and .205 for low seventh. Although the last named coefficient of .205 is very low, the average of all the coefficients is still .51. The low correlation of the low seventh group might in part be accounted for by the fact that there are five more

girls than boys in this group, while in the other three groups there is only one more girl than boys. Girls often make high class grades because of their ability to apply themselves, although their general intelligence may not be so high as the boys', in spite of the latter's lower class grades.

The mean variation of each group shows an interesting comparison. The high and the low eighth subjects have almost an equal variation: 49.2 and 49.7 respectively, whereas the smaller variations of high and low seventh with 39.8 and 41.5 show a greater range. This would also tend to show that the older eighth grade pupils show greater variation than the younger seventh grade pupils.

An examination of the relationship between vocabulary index and chronological age indicates in general that the younger pupils in each group have the highest V. I. and the older students just the opposite. This is a logical conclusion, for the duller children usually start to school later in life, and the younger children in a grade are nearly always the brightest ones.

Table VI indicates the influence of sex upon the size of vocabulary. In each of the four groups the boys have the higher average indices. It was found, however, that in two classes a girl has the highest, and in two, the lowest index, while in two classes a boy has the highest index and in two, the lowest.

A comparison of the two sex groups shows that there is a much larger mean variation among the boys in three of the groups while the girls have a large variation in only one group. Perhaps this would be an indication of the premise that there are more geniuses among men, but also more men who are below the average.

The vocabularies listed in Table V as the average vocabularies for each of the four groups tested were found by multiplying the average vocabulary index of each group by 250. These results show much larger vocabularies than those found by investigations with Terman's,⁴ Kirkpatrick's,³ and other tests. As previously explained the difference is largely due to the size of the dictionary used as a basis for the word list.

The results offered in Table VII, (a), show that the proportional number of errors made in each scholastic group is practically identical. Therefore the thesis that more errors are made in the lower classes is unfounded and the tests may be used in the seventh as well as the eighth, and perhaps in the sixth grade. However, as was to be expected, the lower third of each class had a greater proportion of errors than the upper third. It is interesting to note that, in general, boys make a larger proportion of errors than girls. This conclusion was to be

expected, for it is generally conceded that boys are more likely to take a chance than girls.

This investigation corroborates the evidence brought out by Gerlach's experiments that the False Definition Test would be a valuable aid to the teacher or administrator in determining promotion or school placement. It emphasizes the value of a large vocabulary and points out a definite problem to the schools: how to acquire the greatest number of words which at the same time may be used with maximum ease and effectiveness.

TABLE I

HIGH EIGHTH PUPILS

	V. I.	Av. G.	Age
Total	8236	78	552
Av.	211.2	2	14.1
Median	218.	3	15.
M. V.	49.28
Correlation of V. I. and Av. G.	: .48		

TABLE II

LOW EIGHTH PUPILS

	V. I.	Av. G.	Age
Total	7516	80	519
Av.	192.7	2.05	13.3
Median	191.	3	12.
M. V.	49.78
Correlation of V. I. and Av. G.	: .54		

TABLE III

HIGH SEVENTH PUPILS

	V. I.	Av. G.	Age
Total	6263	74	492
Av.	164.8	1.9	12.9
Median	160.5	1.5	12.5
M. V.	39.848
Correlation of V. I. and Av. G.	: .81		

TABLE IV

LOW SEVENTH PUPILS

	V. I.	Av. G.	Age
Total	5944	73	478
Av.	152.4	1.8	12.2
Median	160.	2.	13.
M. V.	41.579
Correlation of V. I. and Av. G.	: .205		

TABLE V

AVERAGE VOCABULARY IN RELATION TO SCHOLASTIC STATUS

Scholastic status	Vocabulary
High Eighth	52,775
Low Eighth	48,175
High Seventh	41,200
Low Seventh	38,100

TABLE VI

SEX INFLUENCE UPON SIZE OF VOCABULARY

	V. I. (Male)	V. I. (Female)
High Eighth	212.8	209.6
Low Eighth	207.3	178.8
High Seventh	175.3	155.5
Low Seventh	156.	149.6
Total	751.4	693.5
Av.	187.8	173.3

TABLE VII

(a) PROPORTIONAL NUMBER OF ERRORS

Scholastic status	Proportion
High Eighth	.33
Low Eighth	.285
High Seventh	.328
Low Seventh	.33

(b) PROPORTIONAL NUMBER OF ERRORS IN UPPER AND LOWER THIRDS

Scholastic status	Upper Third	Lower Third
High Eighth27	.44
Low Eighth14	.31
High Seventh127	.458
Low Seventh285	.49

(c) PROPORTIONAL NUMBER OF ERRORS SEX INFLUENCE

Sex	Proportion
Male	.326
Female	.307

SUMMARY

1. The average V. I.'s as well as the medians indicate a definite relation between vocabulary and scholastic status.

2. Correlation of vocabulary with school records shows a coefficient of .51. In the high seventh group of pupils this coefficient reaches as high as .81, but in low seventh it is only .21.

3. Although the relation between size and vocabulary and chronological age is not constant, in general the average vocabulary of the younger pupils of each grade seems to be the larger.

4. As a whole the boys' vocabularies are larger than the girls'.

5. The proportional number of errors made by each of the four grades is almost constant.

BIBLIOGRAPHY

- ¹ Bonser, Frederick G., "Vocabulary tests as measures of school efficiency," *School and Society*, vol. 2, pp. 713-718 (November, 1915). New York, Science Press.
- ² Gerlach, Fred M., *Vocabulary Studies*. 1917.
- ³ Kirkpatrick, E. A., "A vocabulary test," *Popular Science Monthly*, vol. 70, pp. 157-164 (February, 1907).
- ⁴ Terman, L. J., "The vocabulary tests as a measure of intelligence," *Jour. of Educational Psychology*, vol. 9, no. 8, pp. 452-456.

TRAINING FOR RAPID READING

BY

J. V. BREITWIESER

One of the most, if not the most, valuable skills that elementary training can give the pupils is the ability of rapid silent reading. Silent reading is a fundamental tool in subsequent educational work. Not only is it necessary for the advanced student to be a good silent reader but it is becoming more and more important for everyone to have this reading ability.

Many investigators have attacked the problem of reading from one or another point of view. Huey's book, *The Psychology and Pedagogy of Reading*, contains most of the data discovered, and remained for a long time, and in many respects still is, one of the most important contributions in the field of silent reading.

In 1921 J. A. O'Brien published his *Silent Reading*. This study presents the important conclusions of investigations up to that time. Most of the studies presented facts concerning factors in the reading process, but failed to offer practical suggestions to teachers that would enable them to increase the speed and accuracy of silent reading.

A comparison of slow readers with fast readers reveals the fact that the chief difficulty is inability to take in or perceive a large number of words at a single glance. The records of eye movements show that the faster rate of reading is accomplished physiologically chiefly by lessening the number of eye fixations to the line and to some extent the shortening of the average duration of the fixations. Any teacher can easily observe the eye jerks or movements of a reader by closely watching the eyes during a silent reading exercise. The eyes are moved by a set of muscles; so, in the last analysis, the proper movement of the eyes depends on muscle activity. Muscle training is therefore an important factor in the problem of increasing reading speed. This is true only when we think of a fairly normal eye. Some pupils will always have to be slow readers because of eye defects. All experiments demonstrate the fact that we must train pupils to see more at a glance or, in the words of O'Brien, to bring about a "more effective utilization of the perceptual span."

The pedagogical problem now becomes one of method. How can the perceptual span be more effectively utilized?

Javal (Emile Javal, "Sur la physiologie de la lecture," *Annales d'Oculistique*, 1878-79) showed that the upper half of the line was the most important in reading: The fixation point moves along between the middle and the top of the small letters. This fact can easily be demonstrated to students by placing a card over the lower halves of words and noting the number that can be identified. When a card is placed over the upper halves of the letters very few words can be identified. When their attention is called to this fact, students can often reduce the amount of eye travel per line.

A second means of training for rapid reading is to create the habit of looking for large units in going over the reading material. This can be done by first presenting short sentences and always having them reproduced as a whole. If the habit of reproducing words is formed there is grave danger of reading by identifying words merely with an eye fixation for each word. As power is gained in recognizing groups of words, these word groups should be increased in length. Always insist that as nearly as possible the whole group be taken in at one glance.

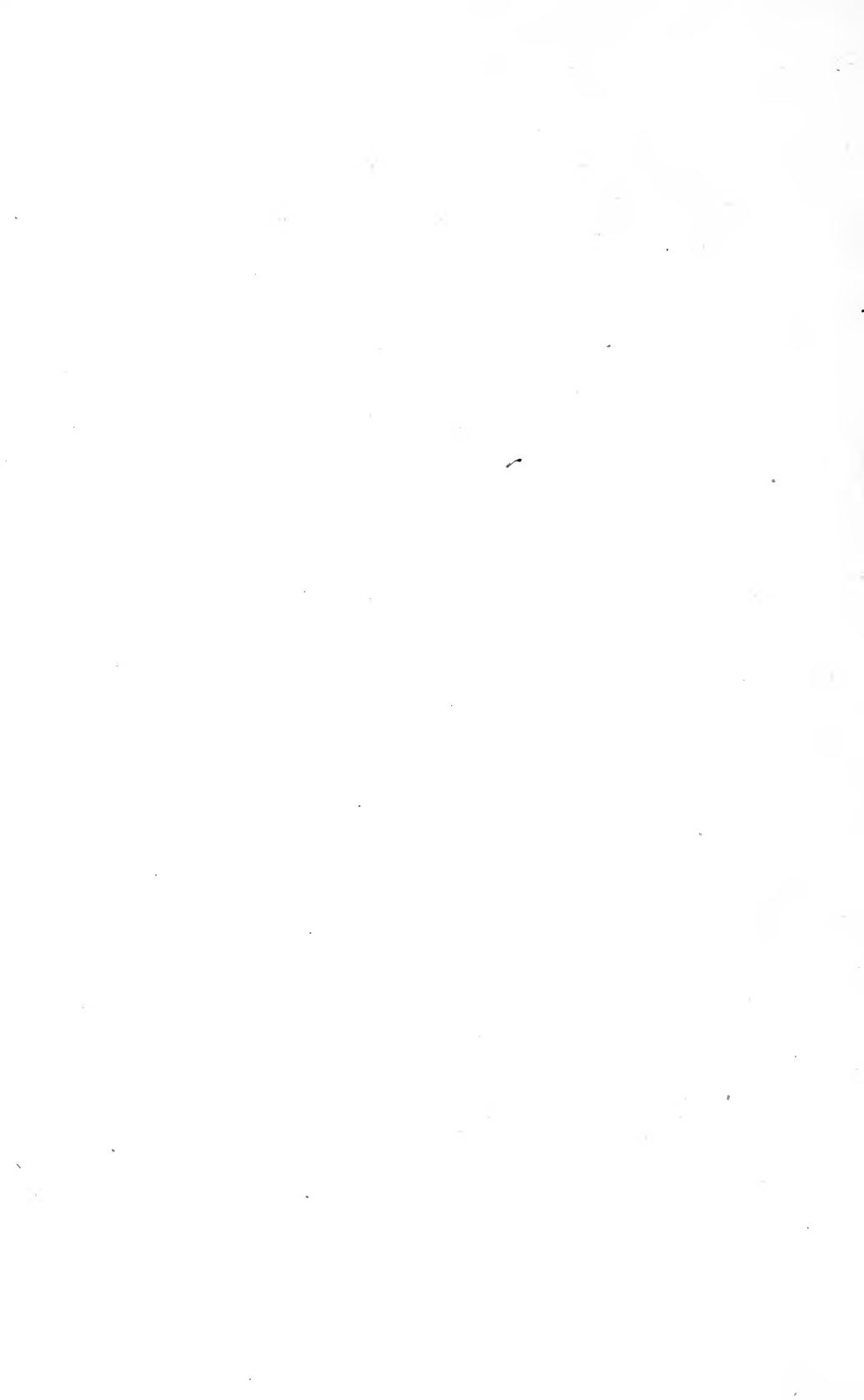
A third factor is the actual visual span. More space on a line can be seen if the book is held farther away from the eyes. Unfortunately a young reader usually holds his book too near to his eyes, and he emphasizes this evil when he comes to a hard word by jerking his book up closely to his eyes. If the word is a long one he can see it as a whole more easily by holding the book farther away. Students should therefore be encouraged to form the habit of holding the book farther away, thus increasing the visual span. This habit will also save the eyes from an early and excessive strain that often leads to myopia or nearsightedness.

Finally, as a fourth point in training for rapid reading, there is the possibility of pacing the eye movements much as we pace a runner or a typist. This point has not been made before in the literature on reading and is based on a series of experiments conducted in 1917 and 1918 and reported in an unpublished Master of Arts thesis by Martin Fereshetian of Colorado College. Mr. Fereshetian, after numerous conferences with the writer, conceived the idea of constructing a piece of apparatus that would expose a page of reading matter, a line at a time, at varying rates. He mounted an endless belt on two pulleys. This belt had slits in it that were just the size of the line of reading matter used. As the belt rotated it exposed the lines successively from top to bottom and from left to right. The pulleys or cylinders were

driven by means of a phonograph motor, the speed of which could be regulated. Subjects were comfortably seated at the usual reading distance from the apparatus, and the belt run at a rate that corresponded to the reading rate. In successive experiments the rate of exposure was increased so as to make it necessary to increase the reading rate if the subject expected to gather all the thought. The reading material was standardized and the subjects constantly questioned as to the content of the exposed pages. The idea was that the subject, if crowded or forced mechanically, would get the feel of more rapid reading, much like the idea behind the "pacing" of a racehorse, a sprinter, or a group of typewriting students.

The results proved this theory to be correct. In sixty cases the average gain in reading time was $33\frac{1}{3}$ per cent. Some slow readers gained as high as 53 per cent, i.e., they were able to read 53 per cent of additional material in the same length of time. Silent reading tests further demonstrated that this faster reading rate was carried over into reading material not exposed in the machine. One college professor remarked, "I never before felt just what it means to read rapidly." It was also noted that students tended to throw their heads back or away from the apparatus when it began to run faster, and that there was a tendency to reduce head and lip movements.

Practically this same pacing effect can be produced by sliding a slotted cardboard screen over a line. A large demonstration card of this kind can be used before classes. A reader getting the "feel" or "set" of rapid reading could then easily carry it over into all his reading. A game of "name the words" may be organized where a phrase or sentence is exposed for one-fifth of a second and the students try to reproduce as many of them as possible—writing them down. The winner is the one who gets the largest number correct. A drop screen can easily be made on which to give the exposures. In this way we can artificially stimulate young readers to a more effective use of that important factor in rapid silent reading, the perceptual span.



A STUDY OF INDIVIDUAL RETESTS

BY

ELISE H. MARTENS

PURPOSE OF STUDY

During the past four years a mental testing program has been carried on in the schools of Oakland, California, which has involved the rather intensive training of approximately 125 individual mental examiners (from the ranks of teachers in the schools) and the giving of some nine thousand individual tests by the Stanford-Binet scale. All the test blanks are filed, as soon as possible after the test has been made, in the central office of the Bureau of Research and Guidance, where a mental test record is kept for every child concerned. In May, 1921, an examination of these record cards revealed the existence of 314 duplicate tests, i.e., two tests made of the same child at different times. In many cases such a retest has been made purposely in the investigation of a definite problem connected with the case; quite as often, however, the pupil concerned has been transferred from one school to another, and the second examination was made by a teacher in the school before it was learned that a test had already been given.

The present study is a discussion of these 314 retests of children in the Oakland schools. Of the 125 examiners who have been at work at one time or another, 84 are involved in the giving of the 628 tests. This group of 84 examiners is a most cosmopolitan one, including those of little or no experience who had very recently begun their training, as well as the more highly trained and expert workers. It is for this reason that the present report is offered as an addition to those which have already been made concerning the constancy of the Intelligence Quotient (I. Q.) as shown by retests. So far as is known to the writer, no one of the investigations thus far carried on has involved so many examiners of such different periods of training. Hence the data at hand will be used to contribute toward the answers to the following questions:

1. With such a large group of comparatively unselected examiners, what degree of correlation exists between retests?
2. Given a common foundation of instructional work in the principles of mental testing, how much do further supervised training and experience contribute toward the agreement between retests?

ANALYSIS OF MATERIALS STUDIED

Six hundred and twenty-eight tests given by a miscellaneous group of 84 examiners during a period of four years may be expected to reveal many differences as to age and intelligence of children tested, as well as the interval of time elapsing between tests. The summary of such differences follows:

Age.—Range at time of first test, 4 years to 16 years. Eighty-three per cent of the children concerned were less than 10 years of age at the time of the first test.

Intelligence.—Range of I. Q. in first test, 33 to 136. Range of I. Q. in second test, 33 to 140. Median I. Q. of each complete set of tests, 86.

Interval of time.—Range, less than 1 month to 3 years. Median interval of time, 13.0 months.

Examiner.—A system of training is in current use in Oakland whereby certification for mental testing is granted a teacher in the department after: (1) a lecture and discussion course in mental testing has been completed; (2) a sufficient number of tests have been submitted for correction to insure the examiner's familiarity with the technique of recording and scoring responses; (3) observation of a test given by the candidate has indicated accurate knowledge of the formulae involved as well as ability to come into satisfactory rapport with the child.

Under such a system of training it is inevitable that numerous test blanks should find their way into the files at the central office which represent work done at a very early stage of the examiner's experience in the actual giving of tests. In such cases indication is always made on the blank accordingly, as a caution against relying too much upon its accuracy. All such tests, however, have been included in this study, with the result that in 50 per cent of the cases one or both of the tests involving the same child were made by an examiner who was still under training, either recently begun or shortly before certification. Without exception, however, the lecture and discussion course in mental testing had been completed or almost completed before actual testing was begun.

The above analysis of data on hand indicates that selective processes have been at work in the aggregation of material contributing to this study. Age preponderance is under 10 years; the median of intelligence is less than 100 I. Q.; 80 per cent of the cases show a time interval between tests of two years or less. In the matter of examiners, however (which is one of the most important items under consideration), there is an approximately equal distribution between the two groups of experienced and inexperienced workers, all having had some previous class work on the general subject of mental testing.

GENERAL AGREEMENT BETWEEN TESTS

In general (except as noted later) test results were taken at their face value as records appeared in the files. A large number of more recent tests had been previously checked in the central office before filing; but a larger number of earlier tests had been filed as submitted by the examiner after training had been completed. Hence inaccuracies may exist, even in the tests of experienced examiners, which have not been discovered.

Considering then the 314 pairs of retests, irrespective of age, intelligence, time interval, or examiner, the following general agreements have been found to exist:

1. Coefficient of correlation (Pearson formula)..... .87
2. Median difference in I. Q. between each pair of tests 6.0 pts.
3. Average difference in I. Q. between each pair of tests..... 7.1 pts.
4. Distribution of tests according to number of points difference
in I. Q.:

No. points difference	Per cent tests	
0-5	45.6	} 76.9
6-10	31.3	
11-15	14.1	
16-20	6.2	
More than 20	2.8	

CHECKING PROCESS INVOLVED

At this stage of the study a checking process was introduced by which all pairs of test blanks, the results of which differed by ten or more points, were carefully analyzed and checked for possible inaccuracies of scoring and age records. Such inaccuracies were handled through a twofold process:

1. Correction of errors wherever discovered.
2. Elimination from further study of 26 pairs of test blanks, one or both of which showed manifest incompleteness or other marked inaccuracies impossible of correction.*

* A "complete" test goes back to a year where every test is passed, and forward until no test within the year is passed. Over 50 per cent of the blanks eliminated stop short with 8 months at the lower limit or 4 (or even 6) months at the upper limit, sometimes both. The remaining tests eliminated revealed lesser degrees of incompleteness *plus* various combinations of errors which were impossible to adjust with anything approaching satisfactory results.

GENERAL AGREEMENT BETWEEN TESTS AFTER CHECKING PROCESS

The remainder of this study, therefore, is concerned with 288 pairs of test blanks, of which those showing a discrepancy of ten or more points in the result have been carefully checked and corrected wherever possible. The following general agreements were now found to exist:

1. Coefficient of correlation (Pearson formula)90†
2. Median difference in I. Q. 5.1 pts.
3. Average difference in I. Q. 5.6 pts.
4. Central tendency of change +0.5 pts.
5. Middle 50 per cent of changes -5.6 to +4.4 pts.
6. Distribution according to number points difference in I. Q.:

No. points difference	Per cent tests	
0-5	53.8	} 88.2
6-10	34.4	
11-15	5.8	
16-20	3.9	} 11.8
More than 20	2.1	

† Cf. also findings of other investigators, given in *Jour. of Ed. Psych.*, Sept. 1921, as follows: Stenquist, .72 (274 cases); Rugg and Colloton, .84 (137 cases); Terman, .93 (435 cases); Cuneo and Terman, .85 (31 cases), .94 (21 cases), .95 (25 cases).

It will be noticed that only in 11.8 per cent of the 288 cases does the difference in I. Q. become more than 10.* A more detailed representation of the distribution of changes is shown in Table I, which indicates the positive and negative differences separately. The central tendency of change is here revealed as +0.5, with a middle 50 per cent range extending from -5.6 to +4.4. Considering the whole number of tests, there seems to be no indication therefore that the result of the second test is apt to be higher or lower than that of the first.

CLASSIFICATION OF TESTS ACCORDING TO AGE

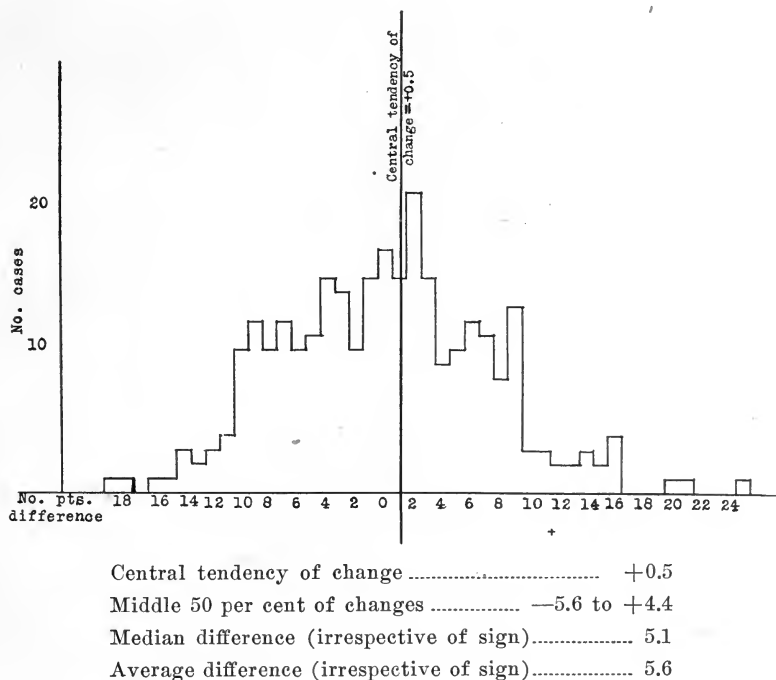
Three age groups were formed, using as a basis the time of the first test. These groups are as follows:

Age	No. cases
†4 yrs. to 7 yrs.....	129
7 yrs. to 10 yrs.....	110
More than 10 yrs.....	49

† The division point of 7 years was made between the first two groups in order that Group 1 might include all or nearly all cases where the first test was made at the time when the child first entered school, thus bringing together the majority of cases where language difficulty, or other obstacles frequently claimed in dealing with the young child, are involved.

* Cf. also *Jour. of Ed. Psych.*, Sept. 1921, in which the following findings are noted of percentage of differences exceeding 10: Garrison, 6.0 per cent (62 cases); Rugg and Colloton, 12.0 per cent (137 cases); Terman, 15.0 per cent (435 cases).

TABLE I
DISTRIBUTION OF CHANGES IN I. Q. BETWEEN FIRST AND SECOND TESTS
(288 cases)



The results obtained through this classification are found in Table II, by which it appears that only by a slight margin do the younger children (up to 7 years of age) show any greater change in the I. Q. of the two tests than do the older pupils. The central tendency of change in either a positive or negative direction is practically negligible with all three groups; notice particularly the change of $+0.1$ with the children above 10 years of age.

CLASSIFICATION ACCORDING TO INTELLIGENCE

Table III shows a redistribution of the 288 cases according to the intelligence indicated by the first test, with the resulting differences revealed in the comparison. One conspicuous fact appears here in our superior group, ranging upward from 110 I. Q., in which there seems to be a tendency to lower that I. Q. in the second test by five points. The significance of this, however, is materially lessened by

the fact that there are only 27 cases in this group. It is still true that, owing to lack of sufficient time for the work, the majority of individual tests made in the years following a child's entrance into school are concerned with problem cases showing more or less real or apparent inability to do the work assigned, and it is rarely the genuinely superior child who presents such a problem for adjustment; hence the small number of retests in the superior group which have been found in the files and thus made available for this study.

TABLE II

COMPARISON OF DIFFERENCES BETWEEN FIRST AND SECOND TESTS, CLASSIFIED
ON THE BASIS OF AGE

Age	4 yrs. to 6 yrs. 11 mos. (129 cases)	7 yrs. to 9 yrs. 11 mos. (110 cases)	10 yrs. or over (49 cases)
Central tendency of change.....	+0.5	-1.2	+0.1
Average change	6.4	5.3	5.3
Median change	6.2	5.2	5.1
Middle 50 per cent of changes	-6.5 to +5.4	-5.1 to +3.2	-4.2 to +5.0

TABLE III

COMPARISON OF DIFFERENCES BETWEEN FIRST AND SECOND TESTS, CLASSIFIED
ON THE BASIS OF INTELLIGENCE

I. Q. (in first test)	Below 90 (171 cases)	90 to 109 (90 cases)	110 or above (27 cases)
Central tendency of change.....	+0.5	-1.3	-5.0
Average change	5.5	6.3	6.8
Median change	4.8	6.4	6.5
Middle 50 per cent of changes	-2.5 to +5.5	-7.5 to +4.5	-8.5 to +1.3

CLASSIFICATION ACCORDING TO TIME INTERVAL

Once more in Table IV is given a redistribution of the cases on hand on the basis of the number of months intervening between the two tests. Again no appreciable difference exists among the three groups, save for the fact that the average change found for those cases in which the time interval ranged from two to three years is 7.6 points; contrast this, however, with the median of 5.6 points which is almost identical with that of the other two groups. The occurrence of several extreme cases within the class of longest time interval accounts for the discrepancy between median and average.

CLASSIFICATION ACCORDING TO EXAMINER

In the distribution made thus far, no distinction has been made between experienced and inexperienced examiners. All were included indiscriminately in the several groups formed. At this point a definite segregation of tests was made, however, into two classes: (1) those pairs, both of which were made by trained and experienced examiners; (2) those pairs, of which one or both were made by examiners who had had only the previous instructional work in the general principles of mental testing and were just beginning the actual giving of tests.

TABLE IV

COMPARISON OF DIFFERENCES BETWEEN FIRST AND SECOND TESTS, CLASSIFIED
ON THE BASIS OF TIME INTERVAL

Time interval	8 days to 12 mos. (145 cases)	13 mos. to 24 mos. (103 cases)	25 mos. to 36 mos. (40 cases)
Central tendency of change.....	+1.2	-0.7	-2.0
Average change	5.3	5.7	7.6
Median change	5.5	5.7	5.6
Middle 50 per cent of changes	-4.2 to +5.2	-6.0 to +4.0	-6.0 to +5.0

TABLE V

COMPARISON OF DIFFERENCES BETWEEN FIRST AND SECOND TESTS, CLASSIFIED
ON THE BASIS OF EXAMINER'S EXPERIENCE

Examiners	Both experienced (156 cases)	One or both inexperienced (132 cases)
Central tendency of change.....	+0.7	+0.6
Average change	4.8	7.0
Median change	3.7	6.2
Middle 50 per cent of changes.....	-5.2 to +4.3	-8.0 to +6.5

Table V shows the results. For the examiners who have had sufficient experience to credit them with certification for the work, the average and median differences are as low as 4.8 and 3.7 points respectively, changes which are increased by 50 per cent or more in those cases where one or both examiners were still under training. The range of the middle 50 per cent of differences shows a similar disadvantage to the inexperienced examiner. Such a comparison can point to only one conclusion, i.e., in order to secure the most satisfactory results, there is a definite need for carefully supervised experience before the examiner is sent forth into the field for general testing work.

ANALYSIS OF CASES DIFFERING BY MORE THAN TEN POINTS

Thirty-four cases, or 11.8 per cent of the whole number studied, show a difference in the I. Q. of the first and second tests of more than ten points. The largest positive difference is 25, and the largest negative change is 19, with a median of 14.5 points difference in the whole series of thirty-four tests. Several questions are important in analyzing this group, for example:

1. Do these large differences occur more frequently in any one age group than in another?
2. To what extent does a language difficulty appear to have any influence on the result of the first test?
3. Does the group include any psychopathic cases?
4. Did both tests fully explore the child's mentality?
5. Is there any apparent relation between large differences in I. Q. and the time interval between tests?
6. Did both examiners have a fair amount of experience?

TABLE VI

ANALYSIS OF THIRTY-FOUR CASES SHOWING DIFFERENCE BETWEEN FIRST AND SECOND TESTS OF MORE THAN TEN POINTS

Age	4 yrs. to 6 yrs. 11 mos. (129 cases)	7 yrs. to 9 yrs. 11 mos. (110 cases)	10 yrs. or more (49 cases)	Totals (288 cases)
Foreign—language difficulty	13	4	0	17
Psychopathic cases	3	0	0	3
One or both tests not fully complete*	2	2	2	6
Doubtful	5	1	2	8
Total	23	7	4	34
Per cent of group..	17.7	6.4	8.1	11.8

Range of time interval, 4 mos. to 33 mos.

Median time interval, 13.5 mos.

Examiners { Both experienced: 13 cases.
 { One or both inexperienced: 21 cases.

* Of the six tests in this group it should be said that they did not show sufficient incompleteness to justify their elimination from the study with those mentioned earlier in the report; a careful analysis of both tests, however, revealed a very possible failure in one or both to explore completely the child's mentality at one or the other end of the scale; hence they are listed here as a separate class.

In Table VI are compiled data indicating the answers to these questions. Twenty-three of the thirty-four cases occur in the youngest age group; moreover, thirteen of these showed a distinct language difficulty when the first test was given at the time of entering school, and thus were not able to reveal the full extent of their mentality until they had received definite help in thought expression through school experience. In the same age group (considering the total number of 129 cases), however, thirty children are included who likewise came from foreign homes with a distinct foreign influence, yet with two tests showing a difference in each case of not more than 9 points and a median change of 5.0. The indications are therefore that the child from the foreign home more often than not in a cosmopolitan community has acquired sufficient knowledge of English by the time he enters school to enable him to make a fairly representative estimate of his intelligence by means of the Stanford-Binet scale.

As to time interval involved, there is no indication that the larger differences accompany the longer periods intervening between tests, since the range of time in this group is from 4 to 33 months, with a median of 13.5 months.

Again some significance may be attached to the fact that of the 34 cases showing large discrepancies, 21 involve tests of which one or both were made by inexperienced examiners. The coupling of this fact with the other considerations of language difficulty, psychopathic conditions, and failure to explore fully the child's mentality is clearly indicative of the reasons for the existence of these 34 cases.

CONCLUSIONS

1. The results of 288 retests by the Stanford revision of the Binet-Simon scale, made by a comparatively unselected group of eighty-four examiners, give a correlation of .90 (Pearson formula). The constancy of the I. Q. therefore appears not to depend upon the specialized methods or personalities of a few highly selected examiners, but to be an objective factor contingent upon the accurate administration of a standardized scale.

2. The constancy of the I. Q. as determined by retests is affected by lack of experience on the part of the examiner. A preliminary course of instruction on the use of the individual mental test plus a sufficient amount of carefully supervised experience to insure the accuracy of administration appear to be necessary factors for the attainment of the most satisfactory results.

3. Only by a slight margin do younger children show any greater change in the I. Q. of the two tests than do the older pupils.

4. The central tendency of change in either a positive or negative direction is in most cases a negligible quantity. There is no general indication that the result of the second test is either higher or lower than that of the first.

5. The degree of intelligence of the pupil tested does not appear to affect the amount of difference between retests. The median change is approximately the same for pupils of superior, normal, and inferior capacity.

6. The interval of time elapsing between tests is not a significant factor in determining the amount of change to be expected. Time intervals of one, two, or three years (or fractions thereof) give the same results in median differences between tests.

7. The greatest care should be taken by the examiner to explore completely the child's mentality, i.e., to go backward to the year in which all tests are passed and forward until every test in a given year is failed in.

8. The factor of language difficulty should be taken into consideration in those schools where this condition is frequently encountered when children enter the kindergarten or first grade. In a cosmopolitan community, however, it is not of sufficiently serious import to lessen the value of a general program of mental testing.

PRINCIPAL REFERENCES*

- ¹ Terman, Lewis M., *The intelligence of school children*, chapter IX, p. 135.
- ² Cuneo, Irene, and Terman, L. M., "Stanford Binet tests of 112 kindergarten children and 77 repeated tests," *Pedagogical Seminary*, 1918, 25, 414-428.
- ³ Garrison, S. C., "Fluctuation of intelligence quotient," *School and Society*, June, 1921.
- ⁴ Poull, Louise E., "Constancy of the I. Q. in mental defectives according to the Stanford revision of Binet tests," *Journal of Educational Psychology*, September, 1921.
- ⁵ Wallin, J. E. Wallace, "The results of retests by means of the Binet scale," *Journal of Educational Psychology*, October, 1921.
- ⁶ Fermon, Marcella L., *Validity of I. Q. as established by retests*, M.A. thesis, Columbia University, May, 1920.
- ⁷ Stenquist, John L., *Unreliability of individual and group intelligence tests in grades 1, 2, and 3*. (Unpublished: includes data of Fermon,⁶)
- ⁸ Colloton and Rugg, "Constancy of I. Q. as determined by retests," *Journal of Educational Psychology*, September, 1921.

* Taken from *The Journal of Educational Psychology*, Sept., 1921.

**RETURN
TO →**

CIRCULATION DEPARTMENT

202 Main Library

642-3403

LOAN PERIOD 1

HOME USE

2

3

4

5

6

ALL BOOKS MAY BE RECALLED AFTER 7 DAYS

1-month loans may be renewed by calling 642-3405

6-month loans may be recharged by bringing books to Circulation Desk

Renewals and recharges may be made 4 days prior to due date

DUE AS STAMPED BELOW

INTERLIBRARY LOAN

JUN 24 1977

UNIV. OF CALIF., BERK.

REC. CIR. III 18 77

FORM NO. DD 6, 40m, 6'76

UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA 94720

Photomount
Pamphlet
Binder
Gaylord Bros.
Makers
Stockton, Calif.
PAT. JAN. 21, 1908

478717

YC 63497

UNIVERSITY OF CALIFORNIA LIBRARY

